Self-calibration of a 1D Projective Camera and its Application to the Self-calibration of a 2D Projective Camera

Olivier Faugeras, Long Quan, Peter Sturm

Self-calibration of a 1D Projective Camera and its Application to the Self-calibration of a 2D Projective Camera

Olivier FAUGERAS[1], Long QUAN[2], and Peter STURM[2]

[1] INRIA Sophia Antipolis, 2004, route des Lucioles – B.P. 93
06902 Sophia Antipolis Cedex, France
Olivier.Faugeras@sophia.inria.fr
http://www.inria.fr/robotvis/personnel/faugeras/faugeras-eng.html
[2] CNRS-GRAVIR-INRIA, ZIRST – 655 avenue de l'Europe
38330 Montbonnot, France
Long.Quan, Peter.Sturm@inrialpes.fr
http://www.inrialpes.fr/movi/people/Quan,Sturm/index.html

Abstract. We introduce the concept of self-calibration of a 1D projective camera from point correspondences, and describe a method for uniquely determining the two internal parameters of a 1D camera based on the trifocal tensor of three 1D images. The method requires the estimation of the trifocal tensor which can be achieved linearly with no approximation unlike the trifocal tensor of 2D images, and solving for the roots of a cubic polynomial in one variable. Interestingly enough, we prove that a 2D camera undergoing a planar motion reduces to a 1D camera. From this observation, we deduce a new method for self-calibrating a 2D camera using planar motions.
Both the self-calibration method for a 1D camera and its applications for 2D camera calibration are demonstrated on real image sequences. Other applications including 2D affine camera self-calibration are also discussed.

1 Introduction

A CCD camera is commonly modeled as a 2D projective device that projects a point in $\mathcal{P}^3$ (the projective space of dimension 3) to a point in $\mathcal{P}^2$. By analogy, we can consider what we call a 1D projective camera which projects a point in $\mathcal{P}^2$ to a point in $\mathcal{P}^1$. This 1D projective camera may seem very abstract, but many imaging systems using laser beams, infra-red or ultra-sounds acting only on a source plane can be modeled this way. What is less obvious, but more interesting for our purpose, is that in some situations, the usual 2D camera model is also closely related to this 1D camera model. One first example might be the case of the 2D affine camera model operating on line segments: The direction vectors of lines in 3D space and in the image correspond to each other via this 1D projective camera model [15, 14]. Other cases will be discussed later.

In this paper, we first introduce the concept of self-calibration of a 1D projective camera by analogy to that of a 2D projective camera which is a very active topic [11, 5, 9, 13, 1] since the pioneering work of [12]. It turns out that the theory of self-calibration of 1D camera is considerably simpler than the corresponding one in 2D. It is essentially determined in a unique way by a linear algorithm using the trifocal tensor of 1D cameras. After establishing this result, we further investigate the relationship between the usual 2D camera and the 1D camera. It turns out that a 2D camera undergoing a planar motion can be reduced to a 1D camera on the trifocal line of the 2D cameras. This remarkable relationship allows us to calibrate a real 2D projective camera using the theory of self-calibration of a 1D camera. The advantage of doing so is evident. Instead of solving complicated Kruppa equations for 2D camera self-calibration, exact linear algorithm can be used for 1D camera self-calibration. The only constraint is that the motion of the 2D camera should be restricted to planar motions. The other applications, including 2D affine camera calibration, are also briefly discussed.

Throughout the paper, vectors are denoted in lower case boldface $\mathbf{x}$, $\mathbf{u}$ ..., matrices and tensors in upper case boldface $\mathbf{A}$, $\mathbf{T}$ ...; Scalars are any plain letters or lower case Greek a, u, A, λ .... The geometric objects are sometimes denoted by plain or Greek letters like l for a 2D line and L for a 3D line whenever it is necessary to distinguish the geometric object l from its coordinate representation by a vector $\mathbf{l}$. Some basic tensor notation is used sometimes: Covariant indices are written as subscripts and contravariant indices as superscripts. e.g. the coordinates of a point $\mathbf{x}$ in $\mathcal{P}^3$ are written with an upper index $\mathbf{x} = (x^1, x^2, x^3, x^4)^T$. A matrix $\mathbf{A}$ is written with two indices like A^i_j, where i indexes rows and j columns. The implicit summation convention is also adopted.

2 1D projective camera and its trifocal tensor

We will first review the one-dimensional camera which was abstracted from the study of the geometry of lines under affine cameras [15, 14]. We can also introduce it directly by analogy to a 2D projective camera.

A 1D projective camera projects a point $\mathbf{x} = (x^1, x^2, x^3)^T$ in $\mathcal{P}^2$ (projective plane) to a point $\mathbf{u} = (u^1, u^2)^T$ in $\mathcal{P}^1$ (projective line). This projection may be described by a 2×3 homogeneous matrix $\mathbf{M}$ as $\lambda\mathbf{u} = \mathbf{M}_{2\times 3}\mathbf{x}$.

We now examine the geometric constraints available for points seen in multiple views similar to the 2D camera case [17, 18, 9, 21, 7]. There is a constraint only in the case of 3 views, as there is no any constraint for 2 views (two projective lines always intersect in a point in a projective plane).

Let the three views of the same point $\mathbf{x}$ be given as follows:

$$\begin{cases} \lambda\mathbf{u} &= \mathbf{M}\mathbf{x}, \\ \lambda'\mathbf{u}' &= \mathbf{M}'\mathbf{x}, \\ \lambda''\mathbf{u}'' &= \mathbf{M}''\mathbf{x}. \end{cases} \tag{1}$$

These can be rewritten in matrix form as

$$\begin{pmatrix} \mathbf{M} & \mathbf{u} & 0 & 0 \\ \mathbf{M}' & 0 & \mathbf{u}' & 0 \\ \mathbf{M}'' & 0 & 0 & \mathbf{u}'' \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ -\lambda \\ -\lambda' \\ -\lambda'' \end{pmatrix} = 0. \tag{2}$$

The vector $(\mathbf{x}, -\lambda, -\lambda', -\lambda'')^T$ cannot be zero, so

$$\begin{vmatrix} \mathbf{M} & \mathbf{u} & 0 & 0 \\ \mathbf{M}' & 0 & \mathbf{u}' & 0 \\ \mathbf{M}'' & 0 & 0 & \mathbf{u}'' \end{vmatrix} = 0. \tag{3}$$

The expansion of this determinant produces a trifocal constraint for the three views

$$T_{ijk} u^i u'^j u''^k = 0, \tag{4}$$

where T_{ijk} is a $2 \times 2 \times 2$ homogeneous tensor whose components T_{ijk} are 3×3 minors (involving all three views) of the 6×3 joint projection matrix by stacking $\mathbf{M}$, $\mathbf{M}'$ and $\mathbf{M}''$.

This trifocal tensor encapsulates exactly the information needed for projective reconstruction in $\mathcal{P}^2$. Namely, it is the unique matching constraint, it minimally parametrizes the three views and it can be estimated linearly. Contrast this to the 2D image case in which the multilinear constraints are algebraically redundant and the linear estimation is only an approximation based on over-parametrization.

Each point correspondence in 3 views $\mathbf{u} \leftrightarrow \mathbf{u}' \leftrightarrow \mathbf{u}''$ yields one homogeneous linear equation for the unknown vector $\mathbf{t}_8$ representing the 8 tensor components T_{ijk} for $i, j, k = 1, 2$:

$$(u^1 u'^1 u''^1, u^1 u'^1 u''^2, u^1 u'^2 u''^1, u^1 u'^2 u''^2, u^2 u'^1 u''^1, u^2 u'^1 u''^2, u^2 u'^2 u''^1, u^2 u'^2 u''^2)\mathbf{t}_8 = 0.$$

With at least 7 point correspondences, we can solve for the tensor components linearly.

A careful normalisation of the measurement matrix is nevertheless necessary just like that stressed in [8] for the linear estimation of the fundamental matrix. The points at each image are first translated so that the centroid of the points is the origin of the image coordinates, then scaled so that the average distance of the points from the origin is $\sqrt{2}$. This is achieved by an affine transformation of the image coordinates in each image: $\bar{\mathbf{u}} = \mathbf{A}\mathbf{u}$, $\bar{\mathbf{u}}' = \mathbf{B}\mathbf{u}'$ and $\bar{\mathbf{u}}'' = \mathbf{C}\mathbf{u}''$.

With these normalised image coordinates, the normalised tensor components $\bar{T}_{ijk}$ are linearly estimated by SVD from $\bar{T}_{ijk} \bar{u}^i \bar{u}'^j \bar{u}''^k = 0$.

The original tensor components T_{ijk} are recovered by de-scaling the normalised tensor $\bar{T}_{ijk}$ as $T_{abc} = \bar{T}_{ijk} A^i_a B^j_b C^k_c$.

3 Self-calibration of a 1D camera from 3 views

The concept of camera self-calibration using only point correspondences became popular in computer vision community following Maybank and Faugeras [12] by solving the so-called Kruppa equations. The basic assumption is that the internal parameters of the camera remain invariant. In the case of the 2D projective camera, the internal calibration (the determination of the 5 internal parameters) is equivalent to the determination of the image ω of the absolute conic.

3.1 The internal parameters of a 1D camera and the circular points

For a 1D camera represented by a 2×3 projection matrix $\mathbf{M}_{2\times 3}$, this projection matrix can always be decomposed into

$$\mathbf{M}_{2\times 3} = \mathbf{K}_{2\times 2} \left(\mathbf{R}_{2\times 2} \; \mathbf{t}_{2\times 1} \right),$$

where

$$\mathbf{K}_{2\times 2} = \begin{pmatrix} \alpha & u_0 \\ 0 & 1 \end{pmatrix}$$

represents the two internal parameters: α the focal length in pixels and u_0 the position of the principal point; the external parameters are represented by a 2×2 rotation matrix $\mathbf{R}_{2\times 2}$,

$$\mathbf{R}_{2\times 2} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and the translation vector $\mathbf{t}_{2\times 1}$.

The object space for a 1D camera is a projective plane, and any rigid motion on the plane leaves the two circular points I and J invariant (a pair of complex conjugate points on the line at infinity) of the plane. Similar to the 2D camera case where the knowledge of the internal parameters is equivalent to that of the image of the absolute conic, the knowledge of the internal parameters of a 1D camera is equivalent to that of the image points $\mathbf{i}$ and $\mathbf{j}$ of the circular points in $\mathcal{P}^2$.

The relationship between the image of the circular points and the internal parameters of the 1D camera follows directly by projecting one of the circular points $I = (i, 1, 0)^T$, where $i = \sqrt{-1}$, by the camera $\mathbf{M}_{2\times 3}$:

$$\lambda \mathbf{i} = e^{-i\theta} \begin{pmatrix} u_0 + i\alpha \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & u_0 \\ 0 & 1 \end{pmatrix} \left(\mathbf{R}_{2\times 2} \; \mathbf{t} \right) \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}.$$

It clearly appears that the real part of the ratio of the homogeneous coordinates of the image of the circular point $\mathbf{i}$ is the position of the principal point u_0 and the imaginary part is the focal length α.

3.2 Determination of the images of the circular points

Our next task is to locate the circular points in the images. Let us consider one of the circular points, say I. This circular point is projected onto $\mathbf{i}$, $\mathbf{i}'$ and $\mathbf{i}''$ in the three views. As they should be invariant because of our assumption that the internal parameters of the camera are constant, we have:

$$\lambda\mathbf{i} = \lambda'\mathbf{i}' = \lambda''\mathbf{i}'' \equiv \mathbf{u},$$

where $\mathbf{u} = (u^1, u^2)^T = \rho(a + ib, 1)^T$ for $\lambda, \lambda', \lambda'', \rho \in \mathcal{C}$.

The triplet of corresponding points $\mathbf{i} \leftrightarrow \mathbf{i}' \leftrightarrow \mathbf{i}''$ satisfies the trilinear constraint (4) as all corresponding points do, therefore,

$$T_{ijk} i^i i'^j i''^k = 0,$$

i.e.

$$T_{ijk} u^i u^j u^k = 0.$$

This yields the following cubic equation in the unknown $x = u^1/u^2$:

$$T_{111}x^3 + (T_{211} + T_{112} + T_{121})x^2 + (T_{212} + T_{221} + T_{122})x + T_{222} = 0. \quad (5)$$

A cubic polynomial in one unknown with real coefficients has in general either three real roots or one real root and a pair of complex conjugate roots. The latter case of one real and a pair of complex conjugates is obviously the case of interest here. In fact, Equation (5) characterizes all the points of the projective plane which have the same coordinates in three views. This is reminiscent of the 3D case where one is interested in the locus of all points in space that project onto the same point in two views (see Section 5.2). The result that we have just obtained is that in the case where the internal parameters of the camera are constant, there are in general three such points: the two circular points which are complex conjugate, and a real point with the following geometric interpretation.

Consider first the case of two views and let us ask the question, what is the set of points such that their images in the two views are the same? This set of points can be called the 2D horopter (h) of the set of two 1D views. Since the two cameras have the same internal parameters we can ignore them and assume that we work with the calibrated pixel coordinates. In that case, a camera can be identified to an orthonormal system of coordinates centered at the optical center, one axis is parallel to the retina, the other one is the optical axis. The two views correspond to each other via a rotation followed by a translation. This can always be described in general as a pure rotation around a point A. A simple computation then shows that the horopter (h) is the circle going through the two optical centers and A, as illustrated in Figure 1.a. In fact it is the circle minus the two optical centers. Note that since all circles go through the circular point (hence their name), they also belong to the horopter curve, as expected.

In the case of three views, the real point, when it exists, must be at the intersection of the horopter (h_{12}) of the first two views and the horopter (h_{23}) of the last two views. The first one is a circle going through the optical centers

C_1 and C_2, the second one is a circle going through the optical centers C_2 and C_3. Those two circles intersect in general at a second point C which is the real point we were discussing, and the third circle (h_{13}) corresponding to the first and third views must also go through this real point C, see Figure 1.b.

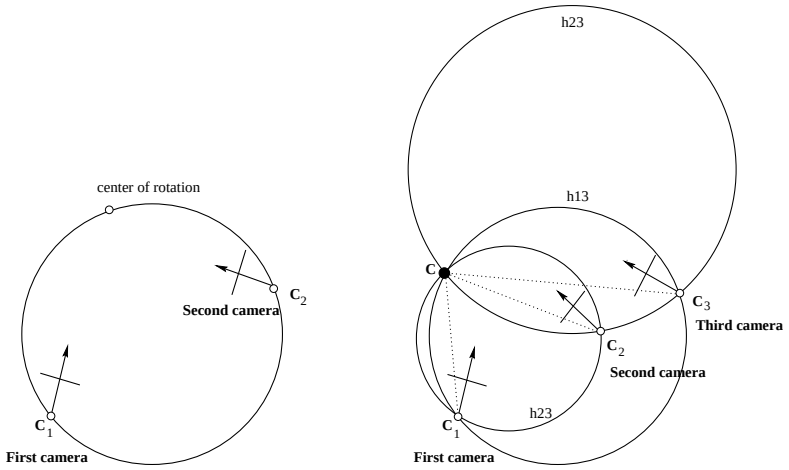

Fig. 1. (a) The two dimensional horopter which is the set of points having the same coordinates in the 2 views (see text). (b) The geometric interpretation of the real point C which has the same images in all three views (see text).

We have therefore established the interesting result that the internal parameters of a 1D camera can be uniquely determined through at least 7 point correspondences in 3 views: the seven points yield the trifocal tensor and Equation (5) yields the internal parameters.

4 Applications

The theory of self-calibration of 1D camera is considerably simpler than the corresponding one in 2D [12] and can be directly used whenever a 1D projective camera model occurs, for instance:

- self-calibration of some active systems using laser beams, infra-reds or ultrasounds whose imaging system is basically reduced to a 1D camera on the source plane;
- self-calibration of a 2D affine camera using line segments,
- partial/full self-calibration of 2D projective camera using planar motions.

The first two types of applications are straightforward. The interesting observation is that the 1D calibration procedure can also be used for self-calibrating a real 2D projective camera if the camera motion is restricted to planar motions. This is discussed in detail in the remaining of this paper.

5 Calibrating a 2D projective camera using planar motions

A planar motion consists of a translation in a plane and a rotation about an axis perpendicular to that plane. Planar motion is often performed by a vehicle moving on the ground, and has been studied for camera self-calibration by Beardsley and Zisserman [3] and Armstrong et al. [1].

Recall that the self-calibration of a 2D projective camera [6, 12] consists of determining the 5 unchanging internal parameters of a 2D camera, represented by a 3×3 upper triangular matrix

$$\mathbf{K} = \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This is mathematically equivalent to the determination of the image of the absolute conic ω, which is a plane conic described by $\mathbf{x}^T(\mathbf{K}^{-1})^T(\mathbf{K}^{-1})\mathbf{x} = 0$ for image point $\mathbf{x}$. Given the image of the absolute conic $\mathbf{x}^T\mathbf{C}\mathbf{x} = 0$, the calibration matrix $\mathbf{K}$ can be found using Cholesky decomposition of the conic matrix $\mathbf{C}$.

5.1 Using known planar motions

By "*known planar motion*", it is meant that the plane of motion is known. The case of unknown planar motions will be treated in the next section.

Without loss of generality and to simplify the explanation, let us assume that the camera is moving on the natural horizontal ground, so the plane of motion is horizontal and the rotation axes are all vertical. We assume a 3D coordinate system with the x- and z-axes on the ground and the y-axis vertical, and a set of point correspondences has been established in 3 images as $(u_i, v_i, 1) \leftrightarrow (u'_i, v'_i, 1) \leftrightarrow (u''_i, v''_i, 1)$.

As the camera is moving on the plane of motion, therefore the trifocal plane—the plane through the camera centers—of the camera is also a plane parallel to the ground and coincident with the plane of motion. Obviously, if restricting the working space to the trifocal plane, we have a perfect 1D projective camera model which projects the points of the trifocal plane onto the trifocal line in the 2D image plane, as the trifocal line is the image of the trifocal plane. In practice, very few or no any points at all really lie on the trifocal plane. However, we may take the orthogonal projection onto the trifocal plane ($y = 0$) of all points $(x_i, y_i, z_i, 1)^T$ in space, *i.e.*

$$(x_i, y_i, z_i, 1)^T \mapsto (x_i, z_i, 1)^T = \mathbf{x}_i.$$

Since the camera plane is vertical and perpendicular to the trifocal plane, this orthogonal projection in the image plane is nothing but taking the horizontal coordinate u_i of the image point $(u_i, v_i, 1)^T$, i.e.

$$(u_i, v_i, 1)^T \mapsto (u_i, 1)^T = \mathbf{u}_i.$$

The points $\mathbf{x}_i = (x_i, z_i, 1)^T$ in $\mathcal{P}^2$ and the points $\mathbf{u}_i = (u_i, 1)^T$ in $\mathcal{P}^1$ are related by the following 1D projective camera model

$$\lambda \begin{pmatrix} u_i \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} x_i \\ z_i \\ 1 \end{pmatrix}.$$

At this point, we have obtained the interesting result that a 1D projective camera model is obtained by considering only the horizontal pixel coordinate of the 2D image points if the camera undergoes a planar motion on the horizontal ground plane!

The 1D self-calibration method just described in the above will allow us to locate the image of the circular points common to all horizontal planes. The circular points are also the intersection points of the line at infinity of the pencil of parallel planes and the absolute conic. We can easily see how the image of the circular points are related to the internal parameters of the 2D camera. The horizontal planes $y = t$ meet the plane at infinity $t = 0$ at the line at infinity whose image is given by $(0, 1, 0)^T \mathbf{K}^{-1}(u, v, 1)^T = 0$, i.e. $v = v_0$. Hence, the location of the trifocal line determines the vertical position of the principal point v_0.

By intersecting the absolute conic given by

$$(u, v, 1)(\mathbf{K}^{-1})^T(\mathbf{K}^{-1})(u, v, 1)^T = 0$$

and the trifocal line, we have

$$\frac{(u - u_0)^2}{\alpha_u^2} + 1 = 0,$$

i.e. the image of the circular points are given by $u_0 \pm i\alpha_u$. It follows that the internal parameters of the 1D camera are exactly two of the internal parameters of the 2D camera: the focal length α_u in horizontal pixels and the horizontal location of the principal point u_0.

If we assume that the image axes are perpendicular and the aspect ratio is known—this is almost always true in practice, this planar motion allows us to calibrate the 2D camera. For a general 2D camera with 5 internal parameters, additional planar motions will be necessary.

5.2 Using unknown planar motions

So far, we have used known planar motions. In the general case we would like a) to be able to determine from three arbitrary images whether they correspond to a planar motion, and b) if it is the case that the motion is planar, to locate the image of the motion plane, then to recover two points on the image of the absolute conic. The determination of the image of the motion plane from fundamental matrices has been reported in [1,3], to which the following paragraph on the planarity test of the motion is therefore closely.

Testing the planarity of the motion

We first estimate the three fundamental matrices of each pair of images which encode the epipolar geometry of each pair. Let $\mathbf{e}_{ij}$, $i \neq j$, $i = 1, 2, 3$ be the epipole in view i with respect to view j. The trifocal lines $\mathbf{t}_i$, $i = 1, 2, 3$ are the lines $\mathbf{e}_{ij} \times \mathbf{e}_{ik}$, $i \neq j \neq k$.

The locus of all points in space that projects onto the same points in two images is the well-known horopter curve (H) [4], in general a twisted cubic. This horopter curve has a very simple definition in terms of the fundamental matrix. Let $\mathbf{F}$ be the fundamental matrix and $\mathbf{x}$ a point in the first image. Its epipolar line in the second image is represented by $\mathbf{Fx}$ and since the image of the 3D point corresponding to $\mathbf{x}$ is the same in the second image, we must have $\mathbf{x}^T\mathbf{Fx} = 0$. Conversely, if $\mathbf{x}$ is a point in the first image such that $\mathbf{x}^T\mathbf{Fx} = 0$, it is the image of a point on the horopter, except if $\mathbf{Fx} = \mathbf{0}$ or $\mathbf{F}^T\mathbf{x} = \mathbf{0}$ because then one has $\mathbf{y}^T\mathbf{Fx} = 0$ or $\mathbf{x}^T\mathbf{Fy} = 0$ for all points $\mathbf{y}$.

The curve of equation $\mathbf{x}^T\mathbf{Fx} = 0$ is a conic (c). The matrix of this conic is $\mathbf{G} = \mathbf{F} + \mathbf{F}^T$ since the antisymmetric part of $\mathbf{F}$ is irrelevant. Note that we have two such identical conics, (c) and (c'), one for each view.

The horopter therefore appears as part of the intersection of two quadratic cones each one being defined by the optical center C (respectively C') and the conic (c) (respectively (c')). The intersection of two quadratic cones is a space curve of degree 4. But note that the points $\mathbf{e}$ and $\mathbf{e}'$ belong to (c) and (c'). This implies that the line (C, C') going through the two optical centers belongs to the curve of intersection since the points on this line project to $\mathbf{e}$ in the first view and to $\mathbf{e}'$ in the second view. But this line is not on the horopter curve since $\mathbf{e}$ satisfies $\mathbf{Fe} = \mathbf{0}$ and $\mathbf{e}'$ satisfies $\mathbf{F}^T\mathbf{e}' = \mathbf{0}$. The remaining chunk is a curve of degree three, in general a twisted cubic. Note that because the images of this cubic in the two views are the conics (c) and (c'), the cubic has to go through the optical centers C and C' [4].

In the case of interest here where we rotate the camera with respect to an axis L, the horopter curve is simpler. Let us consider the images l and l' of L in the two views. They are clearly identical and therefore L belongs to (H). The remaining part must be a conic which can be easily determined.

Consider the plane Π going through the two optical centers C and C' and perpendicular to L. Such a plane is well defined if L does not meet the line (C, C') which we assume to be the case. The intersection of that plane with the two retinal planes determines two 1D cameras with optical centers C and C' and we can look for the points in that plane such that they have the same images in the two 1D cameras (this is the 2D horopter curve (h) for the system of two cameras described in Section 3.2). Let A be the point of intersection of L and Π. We saw in Section 3.2 that (h) is the circle going through C, C' and A.

The conclusion is that in the case of a rotation with respect to an axis L, the horopter curve (H) splits into the line L and the previous circle in the plane Π. Its image (c) (respectively (c')) therefore also splits into two lines, the image line l (respectively the line l') of L and the line p (respectively p') of intersection of Π with the retinal plane.

For a set of three views, we consider the plane Π_{12} corresponding to the first rotation and the plane Π_{23} corresponding to the second. The motion is planar if and only if the two planes Π_{12} and Π_{23} coincide with the trifocal plane.

This gives a test for the planarity of a motion. The three fundamental matrices yield the three trifocal lines l_i represented by $\mathbf{t}_i = \mathbf{e}_{ij} \times \mathbf{e}_{ik}$ $i = 1, 2, 3$, $i \neq j \neq k$. The three matrices $\mathbf{G}_{ij} = \mathbf{F}_{ij} + \mathbf{F}_{ij}^T$, $i \neq j$ (note that $\mathbf{G}_{ij} = \mathbf{G}_{ji}$) define the three conics (c_{ij}) which must split each into two lines, hence the six lines l_{12}, p_{12}, l_{23}, p_{23}, l_{31}, p_{31}. A necessary condition for the motion to be planar is that the six lines l_1, l_2, l_3, and l_{12}, l_{23}, l_{31} are "close" enough. The lines p_{ij} are the image lines of the rotation axes, hence generally different and intersect at a point which is the vanishing point of the direction of the rotation axes (the direction perpendicular to the common plane of motion).

Converting 2D images into 1D images

Now comes the central idea of our method: *the 2D images of a camera undergoing any planar motion reduce to 1D images by projecting the 2D image points onto the trifocal line.*

This can be achieved in at least two ways.

First, if the vanishing point $\mathbf{v}$ of the rotation axes is well-defined. Given a 3D point M with image $\mathbf{m}$, we mentally project it to $\hat{M}$ in the plane of motion, the projection being parallel to the direction of rotation. The image $\hat{\mathbf{m}}$ of this virtual point can be obtained in the image as the intersection of the line $\mathbf{v} \times \mathbf{m}$ with the trifocal line $\mathbf{t}$, i.e. $\hat{\mathbf{m}} = \mathbf{t} \times (\mathbf{v} \times \mathbf{m})$.

Since the vanishing point $\mathbf{v}$ of the axes of rotation and the trifocal line $\mathbf{t}$ are known, this is a well-defined construction, see Figure 2.a.

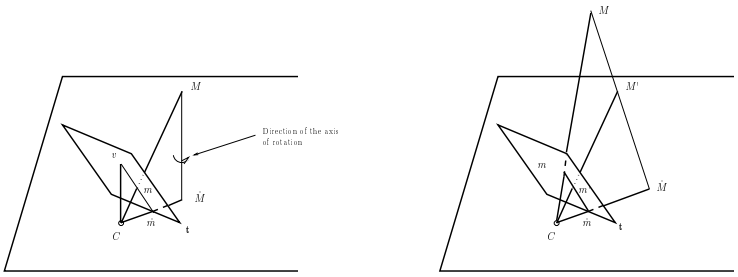

Fig. 2. (a) Creating a 1D image from a 2D image from the vanishing point of the rotation axis and the trifocal line (see text). (b) Creating a 1D image from any pairs of points or any line segments (see text).

Note this is also a projective projection from $\mathcal{P}^2$ (image plane) to $\mathcal{P}^1$ (trifocal line): $\mathbf{m} \mapsto \hat{\mathbf{m}}$.

Alternatively, if the vanishing point does not exist or is poorly defined, we can nonetheless create the virtual points in the trifocal plane. Given two points M and M' with images m and m', the line (M, M') intersects the plane of

motion in $\hat{M}$. The image $\hat{m}$ of this virtual point can be obtained in the image as the intersection of the line (m, m') with the trifocal line $\mathbf{t}$, see Figure 2.b.

Another important consequence of this construction is that *2D image line segments can also be converted into 1D image points!* The construction is even simpler, as the resulting 1D image point is just the intersection of the line segment with the trifocal line.

1D Self-calibration
Once the 2D images are reduced to 1D images, we apply the 1D self-calibration method described in Section 3 to the 1D images.

Estimation of the image of the absolute conic for the 2D camera
Each planar motion generally gives us two points on the absolute conic, together with the vanishing point of the rotation axes as the pole of the trifocal line w.r.t. the absolute conic. The pole/polar relation between the vanishing point of the rotation axes and the trifocal line was introduced in [1]. As a whole, this provides 4 constraints on the absolute conic. Since a conic has 5 d.o.f., at least 2 different planar motions, yielding 8 linear constraints on the absolute conic (see Figure 3.a), will be sufficient to determine the full set of 5 internal parameters of a general 2D camera by fitting a general conic $\mathbf{x}^T\mathbf{C}\mathbf{x} = 0$.

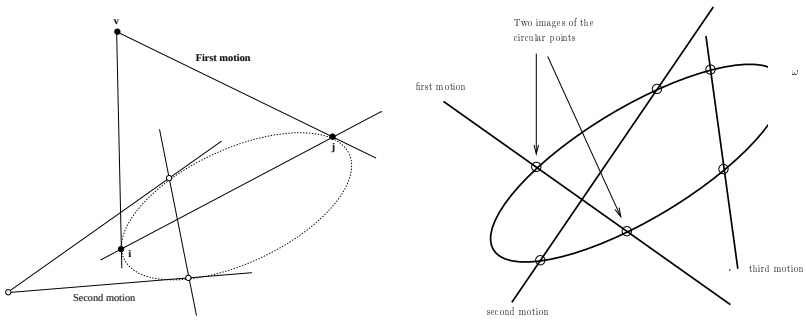

Fig. 3. (a) Computing the image of the absolute conic from two planar motions using the vanishing point of the rotation axes. (b) Computing the image of the absolute conic from three planar motions using only points lying on the conic.

If we assume a 4-parameter model for camera calibration with no image skew (i.e. $s = 0$), one planar motion yielding 4 constraints to determine the 4 internal parameters of the 2D camera is generally sufficient. However this is not true for the very common planar motions such as purely horizontal or vertical motions with the image plane perpendicular to the motion plane (for instance, the planar motion described in Section 5.1 typically belongs to this situation)! It can be easily proven that there are only 3 instead of 4 independent constraints on the absolute conic in these configurations. We need at least 2 different planar motions for determining the 4 internal parameters.

This also suggests that even the planar motion is not purely horizontal or vertical, but close to be, the vanishing point of the rotation axes only constrains loosely the absolute conic. Using only the circular points located on the absolute conic is preferable and numerically stable, but we may need at least 3 planar motions to determine the 5 internal parameters of the 2D camera (see Figure 3.b). Note that the numerical unstability of the vanishing point for nearly horizontal trifocal line was already noticed by Armstrong in [2].

Obviously, if we work with a 3-parameter model for camera calibration with known aspect ratio and no image skew, one planar motion is sufficient as was illustrated in Section 5.1 and [1].

Summary and comparison
We can summarize our algorithm for self-calibrating a 2D camera as follows:

1. Take three views of a scene.
2. Estimate the three fundamental matrices [23, 20].
3. Verify that the motion is planar as described above. If it is not planar, stop.
4. Project the point and line correspondences in the retinal plane using either one of the two methods proposed above.
5. Estimate linearly the trifocal tensor of the 3 corresponding 1D images.
6. Solve for the three roots of (5). This yields two points on the image of the absolute conic.
7. If the number of constraints on the absolute conic is less than the number of internal parameters, go to step 1.
8. Fit a complex ellipse using all available constraints.

As we have mentioned at the beginning of this section the method described in this section is related to the work of Armstrong et al. [1], but there are some important differences which we explain now.

First, our approach gives an elegant insight of the intricate relationship between 2D and 1D cameras for a special kind of motion called planar motion.

Second, it allows us to use only the fundamental matrices of the 2D images and the trifocal tensor of 1D images to self-calibrate the camera instead of the trifocal tensor of 2D images. It is now well known that fundamental matrices can be very efficiently and robustly estimated [23, 20]. The same is true of the estimation of the 1D trifocal tensor [14] which is a *linear* process. Armstrong et al., on the other hand, use the trifocal tensor of 2D images which so far has been hard to estimate due to complicated algebraic constraints. Also, the trifocal tensor of 2D images takes a special form in the planar motion case [1] and the new constraints have to be included in the estimation process.

It may be worth mentioning that in the case of interest here, planar motion of the cameras, the Kruppa equations become degenerate [22] and do not allow to recover the internal parameters. Since it is known that the trifocal tensor of 2D images is algebraically equivalent to the three fundamental matrices plus the restriction of the trifocal tensor to the trifocal plane [16, 10], our method can be seen as an unexpensive way of estimating the full trifocal tensor of 2D images:

first estimate the three fundamental matrices (nonlinear but simple and well understood), then estimate the trifocal tensor in the trifocal plane (linear).

Although it looks superficially that both the 1D and 2D trifocal tensors can be estimated linearly with at least 7 image correspondences, this is misleading since the estimation of the 1D trifocal tensor is exactly linear for 7 d.o.f. whereas the linear estimation of the 2D trifocal tensor is only a rough approximation based on a set of 26 auxiliary parameters for its 18 d.o.f. and obtained by neglecting 8 complicated algebraic constraints.

Third, but this is a minor point, our method may not require the estimation of the vanishing point of the rotation axes.

6 Critical motions for self-calibration

In this section, we describe the camera motions that may defeat the self-calibration method developed in Section 3. We call these camera motions critical in that they give rise to ambiguous solutions to the location of the circular points, thus leading to ambiguous calibration and ambiguous metric reconstruction. More details and proofs are given in [19]. Here we only summarize the results.

- If the camera center remains fixed, but the camera may rotate arbitrarily, 2D reconstruction is obviously impossible while self-calibration is possible.
- If the camera undergoes pure translations, by analogy to the 2D camera case, affine reconstruction is possible, but self-calibration is impossible.
- If the camera moves on a circle, and is oriented in such a way that all points of the circle have fixed projections in all the views, neither affine reconstruction nor self-calibration are possible.

7 Experimental results

The theoretical results for 1D camera self-calibration and its applications to 2D camera calibration have been implemented and experimented on synthetic and real images.

We first used a regular square grid of 25 points to show the very strong stability of focal length w.r.t. high noise level. The principal point coordinate is set to 200 and the focal length to 400. The image resolution with such a setting of internal parameters is about 500 pixels. The 25 points of the grid are projected onto the three 1D images. Finally the positions of the projected points are perturbed by adding various amounts of uniformly distributed pixel noise. The results are presented in Table 1.

We note the very stable results of the estimated focal length. With noise levels up to ± 10 pixels, the estimation remains extremely near to the true value. The degradation for the estimated positions of the principal point and the fixed point is also graceful, but more sensitive to noise.

For the real case, we consider a scenario of a real camera mounted on a robot's arm. Two sequences of images are acquired by the camera moving in two

Noise	±0	±1	±2	±3	±4	±5	±6	±7	±8	±9	±10
α	400	400.2	400.4	400.7	400.9	401.1	401.3	401.5	401.7	401.8	401.9
u_0	200	205.9	211.8	217.7	223.6	229.5	235.3	241.1	246.8	252.6	258.3
σ	0.0	0.03	0.06	0.09	0.1	0.2	0.3	0.3	0.5	0.6	0.9

Table 1. Table of the estimated internal parameters α and u_0 by 1D self-calibration method from 3 synthetic 1D images with uniform pixel errors of different levels. The fourth row σ shows the standard deviation of the reprojection error for the estimation of the trifocal tensor.

different planes. The first sequence contains 7 (indexed from 16 to 22) images (cf. Figure 4) and the second contains 8 (indexed from 8 to 15).

The calibration grid was used to have the ground truth for the internal camera parameters which have been measured as $\alpha_u = 1534.7$, $\alpha_v = 1539.7$, $u_0 = 281.3$ and $v_0 = 279.0$ using the standard calibration method.

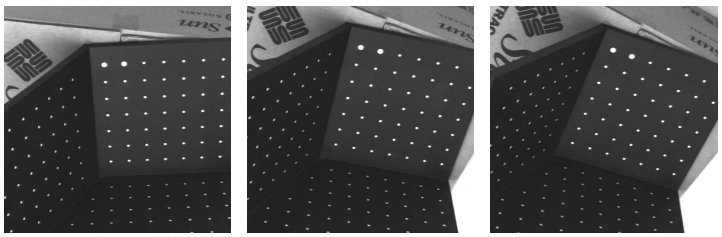

Fig. 4. Three images of the first planar motion.

We take triplets of images from the first sequence and for each triplet we estimate the trifocal line and the vanishing point of the rotation axes by the 3 fundamental matrices of the triplet. The 1D self-calibration is applied for estimating the images of the circular points along the trifocal lines. To evaluate the accuracy of the estimation, the images of the circular points of the trifocal plane are re-computed in the image plane from the known internal parameters by intersecting the image of the absolute conic with the trifocal line. Table 2 shows the results for different triplets of images of the first sequence.

Since we have more than 3 images for the same planar motion of the camera, we could also estimate the trifocal line and the vanishing point of the rotation axes by using all available fundamental matrices of the 7 images of the sequence. The results using redundant images are presented for different triplets in Table 3. We note the slight improvement of the results compared with those presented in Table 2.

The same experiment was carried out for the other sequence of images where the camera underwent a different planar motion. Similar results to the first image

Image triplet	Fixed point	Circular points by self-calibration	Circular points by calibration
$(16, 19, 22)$	493.7	$290.7 \pm i2779.1$	$310.3 \pm i2650.3$
$(16, 20, 22)$	421.8	$250.1 \pm i2146.3$	$273.9 \pm i2153.5$
$(17, 19, 21)$	533.1	$291.3 \pm i2932.4$	$241.3 \pm i2823.1$
$(16, 18, 20)$	617.8	$238.5 \pm i2597.6$	$238.1 \pm i2791.5$
$(18, 20, 22)$	368.3	$230.6 \pm i2208.2$	$272.1 \pm i2126.2$

Table 2. Table of the estimated positions of the images of the circular points by self-calibration with different triplets of images of the first sequence.

Image triplet	Circular points	Fixed point
$(16, 18, 20)$	$245.5 \pm i2490.5$	590.0
$(18, 20, 22)$	$221.4 \pm i2717.8$	384.4
$(16, 20, 22)$	$236.2 \pm i2617.3$	452.9
$(16, 19, 22)$	$240.0 \pm i2693.4$	488.0
$(17, 19, 21)$	$304.7 \pm i2722.7$	516.6
known position by calibration	$262.1 \pm i2590.6$	

Table 3. Table of the estimated positions of the image of circular points with different triplets of images. The trifocal line and the vanishing point of the rotation axes are estimated using 7 images of the sequence instead of the minimum of 3 images.

sequence are obtained, we give only the result for one triplet of images in Table 4 for this sequence.

Image triplet	Fixed point	Circular points by self-calibration	Circular points by calibration
$(8, 11, 15)$	927.2	$269.7 + i1875.5$	$276.5 + i1540.1$

Table 4. Table of the estimated position of the image of circular points with one triplet of second image sequence.

Now two sequences of images each corresponding to a different planar motion yield four distinct imaginary points on the image plane which must be on the image ω of the absolute conic. We therefore fit to those four points an imaginary ellipse using standard techniques and compute the resulting internal parameters. Note that we did not use the pole/polar constraint of the vanishing point of the rotation axes on the absolute conic as it was discussed in Section 5.2 that this constraint is not numerically reliable.

The ultimate goal of self-calibration is to get 3D metric reconstruction. 3D reconstruction from two images of the sequence is performed by using the estimated internal parameters as illustrated in Figure 5. To evaluate the reconstruction quality, we did the same reconstruction using the known internal parameters. Two such reconstructions differ merely by a 3D similarity transformation which could be easily estimated. The resulting normalised relative error of the reconstruction from two images by self-calibration with respect to the

reconstruction by off-line calibration is 3.4 percent. More intensive experimental results could be found in the extended version which can be downloaded from `http://www.inrialpes.fr/movi/people/Quan/publication`.

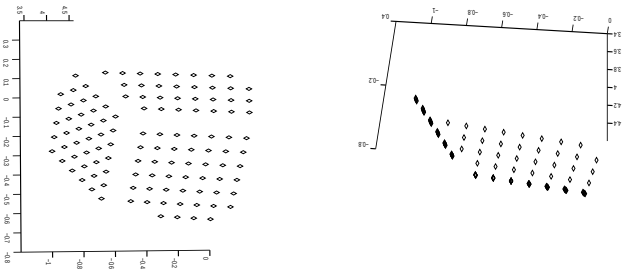

Fig. 5. Two views of the resulting 3D reconstruction by self-calibration.

8 Conclusions and other applications

We have first established that the 2 internal parameters of 1D camera can be uniquely determined through the trifocal tensor of three 1D images. Since the trifocal tensor can be estimated linearly from at least 7 points in three 1D images, the method of the 1D self-calibration is a real linear method (modulo the fact that we have to find the roots of a third degree polynomial in 1 variable), no any over-parameterisation was introduced.

Secondly, we have proven that if a 2D camera undergoes a planar motion, the 2D camera reduces to a 1D camera in the plane of motion. The reduction of a 2D image to a 1D image can be efficiently performed by using only the fundamental matrices of 2D images. Based on this relation between 2D and 1D images, the self-calibration of 1D camera can be applied for self-calibrating a 2D camera. Our experimental results based on real image sequences show the very large stability of the solutions yielded by the 1D self-calibration method and the accurate 3D metric reconstruction that can be obtained from the internal parameters of the 2D camera estimated by the 1D self-calibration method.

Acknowledgement This work has been supported in part by European project CUMULI. We would also like to thank R. Mohr for interesting discussion and the anonymous ECCV reviewers for the constructive comments.

References

1. M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. ECCV, 3–16, 1996.
2. M. Armstrong. Self-calibration from image sequences. Ph.D. Thesis, University of Oxford, 1996.

3. P.A. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. *Europe-China Workshop on GMICV*, 214–221. 1995.
4. T. Buchanan. The twisted cubic and camera calibration. CVGIP, 42(1):130–132, 1988.
5. O. Faugeras. Stratification of three-dimensional vision: Projective, affine and metric representations. JOSA, 12:465–484, 1995.
6. O. Faugeras and S. Maybank. Motion from point matches: Multiplicity of solutions. IJCV, 3(4):225–246, 1990.
7. O. Faugeras and B. Mourrain. About the correspondences of points between n images. *Workshop on Representation of Visual Scenes*, 37–44, 1995.
8. R. Hartley. In defence of the 8-point algorithm. ICCV, 1064–1070, 1995.
9. R.I. Hartley. A linear method for reconstruction from lines and points. ICCV, 882–887, 1995.
10. A. Heyden. *Geometry and Algebra of Multiple Projective Transformations*. Ph.D. thesis, Lund University, 1995.
11. Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. IJCV, 22(3):261–289, 1997.
12. S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. IJCV, 8(2):123–151, 1992.
13. R. Mohr, B. Boufama, and P. Brand. Understanding positioning from multiple images. AI, (78):213–238, 1995.
14. L. Quan. Uncalibrated 1D projective camera and 3D affine reconstruction of lines. CVPR, 60–65, 1997.
15. L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *Trans. PAMI*, 19(8):834–845, 1997.
16. L. Quan. Algebraic Relations among Matching Constraints of Multiple Images. Technical Report INRIA, RR-3345, Jan. 1998 (also TR Lifia-Imag 1995).
17. A. Shashua. Algebraic functions for recognition. *Trans. PAMI*, 17(8):779–789, 1995.
18. M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. *DARPA Image Understanding Workshop*, 271–283, 1990.
19. P. Sturm. Vision 3D non calibrée : contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage. Ph.D. Thesis, INPG, 1997.
20. P.H.S. Torr and A. Zissermann. Performance characterization of fundamental matrix estimation under image degradation. MVA, 9:321–333, 1997.
21. B. Triggs. Matching constraints and the joint image. ICCV, 338–343, 1995.
22. Cyril Zeller and Olivier Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. Research Report 2793, INRIA, February 1996.
23. Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. AI, 78:87–119, 1995.